DOI: 10.1561/9781638284772.ch11

Chapter 11

Image and Video Data Analysis

By Liyue Fan

11.1 Introduction

Image and video data have been increasingly generated and their analysis is ubiquitous in our daily life. The richness of visual data as well as recent technological advances in computer vision inflict great privacy concerns. Classic differential privacy, which originated in statistical databases, has been applied to generating aggregate statistics or training machine learning models, while protecting the privacy of input data. The rigorous nature of DP makes it desirable for protecting sensitive information that can be inferred from image and video data. As a recent research direction, this chapter discusses the role of differential privacy in image and video analysis, especially the developments in sanitizing image and video data. Specifically, we will look at DP notions that aim to quantify sensitive information in image and video data. Furthermore, we will introduce practical privacy and quality measures for evaluating DP methods. We realize that much of the image and video analysis is based on machine learning techniques, and thus direct interested readers to Part II for learning with differential privacy.

As a society, we collectively generate massive amounts of image and video data at a high speed. Studies show that photo-sharing is one of the most common activities of over two-thirds of American adults who now use social media [Dug15; Smi17].

(a) Blurred face of a survivor of human trafficking [UNO10].

(b) Mosaic faces in a police raid [Ait16].

Figure 11.1. Example news images where faces have been obfuscated for privacy (best when zoom).

Approximately 1,000 Terabytes of video data are generated every minute [Kno13] from social media sites to surveillance cameras. Applications that rely on image and video data are ubiquitous, from surveillance to tele-medicine, from facial recognition to eye tracking. In this chapter we will look at the application of differential privacy to image and video data and understand how DP methods intersect with image and video applications.

11.1.1 Perceptions and Expectations for Visual Privacy

Visual anonymity in images and videos is very important for communication research and modern journalism. By providing anonymity, individuals are able to speak more freely [Ano98; Mar01; Sco04], e.g., for conveying sensitive information, expressing marginalized views, and protecting them from retaliation or subsequent contact. Currently visual anonymity is often achieved by blurring or mosaicing faces in pictures and video clips. Figure 11.1 shows two examples where visual anonymity is necessary for those being photographed.

Furthermore, privacy is a significant concern in social media [SCB17] and digital surveillance [WR14]. Researchers have studied the human perceptions of sensitive visual content [LTKC18], shared privacy expectations for online images [Hoy+20], as well as the impact on the viewer's experience when transforming parts of images for enhanced privacy [Li+17; Has+18]. Examples of content categories commonly deemed sensitive in image and video data are identity, children, nudity, and medical condition (e.g., hospital stay).

Moreover, laws and regulations protect the privacy of image and video data. The HIPAA privacy rule requires that full-face photographs and any comparable images must be removed under the Safe Harbor Method. Photographs that can be linked to a patient are considered identifiable information, and therefore, they are subject to HIPAA requirements [NBB19]. The General Data Protection Regulation (GDPR), effective since 2018, aims to increase the privacy of the European Union's

citizens and visitors. It is important for organizations and businesses to consider image and video data as “personal data" or “sensitive personal data" in GDPR terms, and implement and ensure privacy protection accordingly.

11.1.2 Existing Privacy Methods

Various privacy solutions have been proposed for image and video analysis. We categorize commonly adopted image and video privacy methods that do not depend on differential privacy in the following groups.

- **Standard obfuscation.** Popular image obfuscation techniques are *pixelization* (also referred to as mosaicing), *blurring*, and *masking*. The goal is to obscure the content such that it is no longer recognizable. Pixelization can be achieved by superposing a rectangular grid over the original image and averaging the color values of the pixels within each grid cell. On the other hand, blurring, i.e., Gaussian blur, removes details from an image by convolving the 2D Gaussian distribution function with the image. Social media platforms may provide their own implementations, e.g., YouTube face blur [SP17] for video uploads. Masking replaces sensitive content with uninformative pixel values, e.g., a solid rectangle of black pixels over a face.
- **Fusion and perturbation.** Image *fusion* and *perturbation* have also been adopted for visual privacy. For instance, Newton et al. [NSM05] proposed to achieve k-anonymity for a set of face images. Their method, named k-same, “averages” face data for a group of individuals, such that each face in the published dataset appears at least k times. In [OR15], a face “morphing" scheme was proposed where the input face image is mixed with another face image to suppress gender information. Recent works, such as [MRR18; RGRB19], show the promise of adversarial perturbations. In [MRR18], perturbed face images could confound gender classifiers, while preserving the accuracy of face matchers. In [RGRB19], adversarial images were created by using a fast flipping attribute technique, and were able to fool DNNs networks in predicting binary facial attributes.
- **Cryptography.** A number of cryptography-based solutions have been developed to utilize untrusted service providers for image storage, sharing, and analysis. For instance, P3 [RGO13] enables privacy-preserving image sharing by encrypting the significant DCT coefficients, and authorized recipients can decrypt and reconstruct the input image. Furthermore, several approaches have been proposed to perform analysis on encrypted image data, e.g., for privacy-preserving image retrieval [Xia+16], extracting features [HLP12], and learning models [BCCW19].

11.1.3 Privacy Risks

Here we review a range of privacy risks associated with sharing image and video data. An in-depth discussion on privacy inference attacks in machine learning models is available in Chapter 5.

Re-identification

Given obfuscated image data, re-identification attacks aim to predict the identity of individuals or the class label of objects. For example, McPherson et al. [MSS16] developed neural network-based models which could be trained to re-identify faces and hand-written digits, and to recognize objects, on images obfuscated with pixelization, YouTube face blur [SP17], and P3 [RGO13] image sharing system. The re-identification rate of faces is up to 98%, after performing pixelization with a relatively large window, e.g., 16×16 pixels. Similarly, Hill et al. [HZSS16] showed that text obfuscated by pixelization can be reconstructed with a large accuracy using hidden Markov models (HMM). Those attack results show that existing image obfuscation techniques may yield unrecognizable images by human users, but state-of-the-art image recognition algorithms, e.g., deep learning based techniques, can successfully recover sensitive information.

Attribute Inference

In biometric template matching, visual data is utilized for recognition. Attribute inference refers to the estimation of other personal attributes, such as gender, age, and facial expression. Dantcheva et al. [DER15] surveyed techniques to extract a range of soft biometrics, such as face, body, fingerprint, hand, and iris, from image and video data, and recent results on the accurate estimation of demographic attributes (e.g., age, gender, race and ethnicity) and medical attributes (e.g., health and body weight). Wang and Kosinski [WK18] showed that sexual orientation can be inferred from facial images, with an accuracy of around 83% to 91%.

Model Inversion

When machine learning models are shared, sensitive training data may be reconstructed via model inversion attacks. Image analysis models have been targeted in model inversion. Fredrikson et al. [FJR15] reconstructed face images from trained neural network models for facial recognition. Zhang et al. [Zha+20] proposed several techniques to reconstruct images with higher quality. As seen in Figure 11.2, the reconstructed image needs not be perfect in order for human or algorithms to recognize the identity information. It is reported that human users can identify the reconstructed faces with an average of 80% accuracy [FJR15], and using state-of-the-art classifiers the reconstructed faces can be identified with accuracy of up to 82% [Zha+20].

Figure 11.2. Example training image (a) and reconstructed image (b) given an identity label and a trained facial recognition model [FJR15].

Membership Inference

Membership inference tells whether a given record is present in the input dataset which is used to produce certain outputs, e.g., aggregate statistics, recommender systems, and machine learning models. Recent work studied membership inference attacks on image models [SSSS17; HRSF20]. An adversary is able to learn whether an image is part of the training set that produced a given model, by exploiting overfitting artifacts on training data.

11.1.4 Application of Differential Privacy

Differential privacy (DP) has become the state-of-the-art privacy paradigm for statistical databases. As introduced in Chapter 1, in *central* DP model, a trusted server is responsible for data aggregation and analysis, and the presence of any record in the input is protected. In the *local* DP model (see Chapter 2), the server is no longer trusted, and the exact value of each input record is protected. There are two general approaches for applying DP to image and video analysis as follows.

- **Training Machine Learning Models.** In image and video analysis, DP can be applied to training models while protecting the presence of each training sample. For instance, Abadi et al. [Aba+16] proposed differentially private stochastic gradient descent (DP-SGD) for deep learning and the moment accountant (MA) technique to account for differential privacy across training epochs. With training data are distributed at different sites, DP can be achieved in a federated learning setup [Li+19], or via private aggregation of teacher ensembles [Pap+16]. Recently, DP has also been applied to train generative adversarial networks to produce synthetic images [Xie+18; TKP19].
- **Sanitizing Image and Video Data.** A different approach is to apply DP to sanitizing sensitive information in image and video data, and the sanitized data can be shared with untrusted parties for further analysis. Figure 11.3

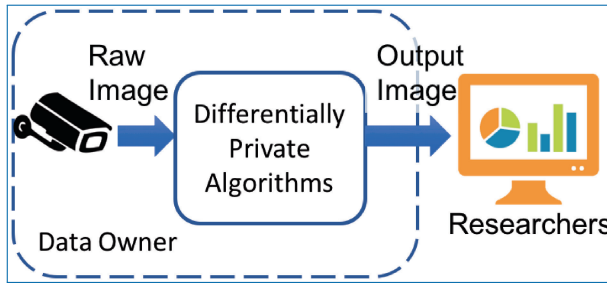


Figure 11.3. Differential Privacy Setting for Sanitizing Image and Video Data. Note that a video is a sequence of images, also known as video frames.

depicts such a setting where a data owner wishes to share image or video data with untrusted recipients, e.g., researchers, servers, or the greater public. The data owner must sanitize the data prior to its publication, in order to guarantee DP. Thanks to the resistance to post-processing [DR+14], any analysis performed on the sanitized data would not inflict additional DP cost.

In-depth analysis of DP for machine learning, including its applications and challenges, is conducted in other chapters of the book. Therefore, in this chapter we primarily discuss the second approach, where DP is applied to *sanitizing image and video data*. We identify two main advantages of the sanitization approach: (1) it offers resistance to post-processing, as mentioned above; (2) it is compatible with personalized privacy, e.g., user-specified ε and δ values for DP guarantees. Those properties would make DP highly desirable for conducting privacy-preserving analysis and for meeting users' privacy needs.

Overview of the Chapter

In the rest of the chapter, we will introduce challenges of applying DP to image and video data sanitization and review the progress of the research community to this date (Section 11.2). In addition to theories, in Section 11.3 we will also introduce important practical privacy protections and utility measures that methods with DP guarantees should keep in mind. Finally, in Section 11.4, we will discuss challenges should be addressed in collaboration with other communities, such as understanding the user perceptions and the deployment in real systems.

11.2 Sanitizing Image and Video Data with DP

In this section, we introduce challenges of applying DP to image and video data sanitization and review the progress made by the research community. Recall our problem setting in Figure 11.3. The sanitization algorithm operates on image or

video data, such that the output does not allow an adversary (e.g., researchers or the greater public) to infer much about sensitive information in the input data.

The key question to address by DP methods in image and video settings is: *what information is protected by DP?* Not surprisingly, this question also helps us categorize and comparatively analyze recent developments in DP for image/video data made by the research community. With this question in mind, we introduce challenges and solutions for sanitizing image and video data with DP. Recall that a video is a sequence of images, i.e., video frames. We will start our discussion with image sanitization, i.e., obfuscation.

11.2.1 Pixel-Level Privacy for Images

For simplicity, let's consider inputs to DP algorithms are grey-scale images. A grey-scale image is a matrix and elements in the matrix are integer pixel values between 0 and 255 (i.e., 0 is black and 255 is white).

Protecting a Single Pixel

In order to adapt differential privacy to image data, a straight-forward idea is to consider pixels as "records" of a database. Therefore, a DP algorithm should protect the values of individual pixels. In a recent study [Joh+20], the following notion was proposed to protect each pixel.

Definition 11.1. [Pixel-DP] Let s denote a randomized algorithm and S be any subset of the image space of s. Then, we say s is (ε, δ)-differentially private if for any S and any pair of neighboring inputs x and x',

$$\Pr[s(x) \in S] \leq e^{\varepsilon} \Pr[s(x') \in S] + \delta \tag{11.1}$$

where neighboring inputs x and x' correspond to two images that differ by at most one pixel.

As can be seen, the above definition is very similar to the classic (ε, δ)-DP [DR+14]. Two input images are *neighbors* if they differ by at most one pixel. The authors of [Joh+20] adopted a randomized mechanism for each pixel such that either the real pixel or a default value, e.g., 127, is reported with a coin flip. Applied to images taken by eye tracking devices, the authors argued that per pixel randomization is more suitable for pupil tracking applications, compared to blur-based obfuscations.

Protecting Multiple Pixels Simultaneously

An immediate concern regarding Definition 11.1 is that hiding the presence of one pixel in the input image may not provide strong privacy, i.e., hiding sensitive information in the input image. Let's look at some images based on widely used PETS

(a) One Sample PETS Image (b) An Example Neighbor Image

Figure 11.4. (a): sample image from PETS dataset where each red rectangle represents some sensitive information and contains ~ 360 pixels. (b): an example neighboring image for (a) by removing the leftmost pedestrian. [Fan18]

dataset [Lea+15] as an example. This dataset contains video frame sequences widely used in Multiple Object Tracking studies. Each red rectangle in Figure 11.4(a) illustrates one type of sensitive information, such as a pedestrian, a van, an object on grass, and a signage; and each rectangle contains $\sim$ 360 pixels, i.e., much higher than one.

We are thus motivated to consider a stronger privacy model, in which sensitive information represented by *multiple* pixels should be protected. [Fan18] proposed a customizable notion for neighboring images.

Definition 11.2. [m-Neighborhood] Two images I_1 and I_2 are m-neighboring images if they have the same dimension and differ by at most m pixels.

As can be seen, Definition 11.2 is a generalization of the 1-pixel neighborhood discussed above. In comparison, Definition 11.2 provides *stronger* privacy: allowing up to m pixels to differ enables us to protect the presence or absence of any sensitive information which can be represented by those pixels in an image. Recall object, text, or person in Figure 11.4. The following is a strict ε-DP definition for m-neighboring images proposed in [Fan18].

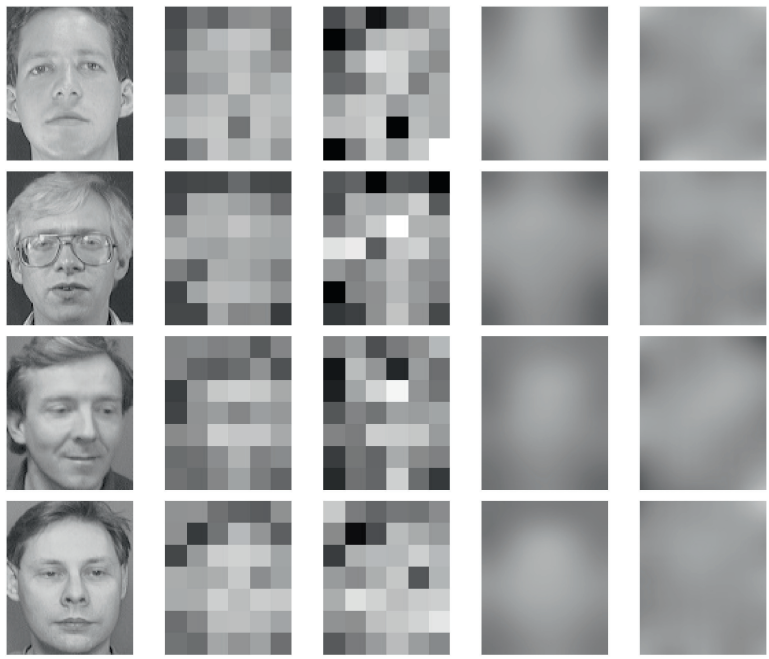
Definition 11.3. [Image-DP] A randomized mechanism $\mathcal{A}$ gives ε-differential privacy if for any m-neighboring inputs I_1 and I_2, and for any possible output $\widetilde{I} \in \mathit{Range}(\mathcal{A})$,

$$\Pr[\mathcal{A}(I_1) = \widetilde{I}] \leq e^{\varepsilon} \Pr[\mathcal{A}(I_2) = \widetilde{I}] \tag{11.2}$$

where the probability is taken over the randomness of $\mathcal{A}$.

With Image-DP, an adversary cannot distinguish between any pair of neighboring images by observing the output image. The privacy of the pedestrian, and any other sensitive information represented by at most m pixels, can thus be protected.

Table 11.1. Qualitative Comparisons for Differentially Private Image Obfuscation: each column represents one obfuscation method; each row lists the obfuscation outcomes for the same input image from the *AT&T* faces dataset. We observe that DP obfuscations inflict only a small quality loss, compared to non-private obfuscations.



When adopting the definitions above, a data owner can choose an appropriate m value in order to customize the level of privacy protection, i.e., achieving indistinguishability in a smaller or larger range of neighboring images. Note that it is assumed that removing those pixels is sufficient to protect the privacy of the underlying information, by definition of differential privacy [DR+14].

Applied to Standard Obfuscation

Differentially private obfuscation mechanisms for *pixelization* and *Gaussian blur* were proposed in [Fan18; Fan19a]. Sample *AT&T* images are provided in Table 11.1. For pixelization (Pix and DP-Pix), the block size is set to 16×16; for Gaussian blur (Blur and DP-Blur), the kernel size is set to 99×99. The image DP parameters are set as $m = 16$ and $\varepsilon = 0.5$ for both pixelization and Gaussian blur. Applied to image obfuscation, Image DP inflicts a small quality Loss, compared to non-private obfuscations. Quantitative quality as well as privacy can be measured for DP image obfuscations, which are discussed in depth in the next section.

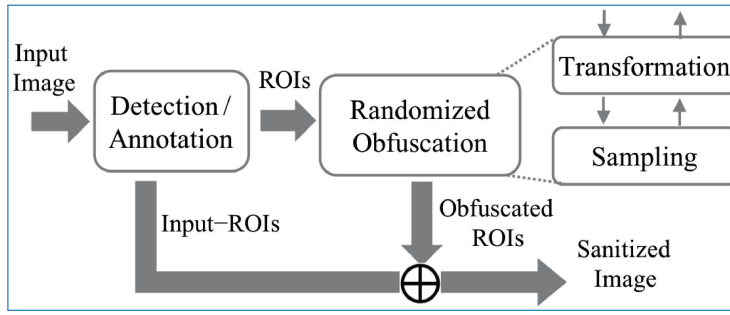


Figure 11.5. A Privacy-Preserving Image Sharing Framework with Perceptual Obfuscation [Fan19b].

Extension to Multi-channel Images

Considering image data with multiple channels, such as RGB (red-green-blue) and HSV (hue-saturation-value) images, each channel may not be independent of the other channels. A straight-forward extension of image DP is to split the privacy budget across multiple channels and apply DP methods accordingly.

11.2.2 Perceptual Image Privacy

In the previous subsection, we quantified image privacy directly with pixels, which is an intuitive approach to extend the classic DP notion to image data. However, the way an image conveys its content is unique and complex. Therefore, we argue that it could be beneficial to first quantify the perceived information from an input image and then apply rigorous privacy. In fact, certain image modifications that inflict pixel value changes may not significantly affect the human perception of image content, for instance, after JPEG image compression [PM92] or adding a small constant to every pixel. The *challenge* is thus to effectively model what can be perceived in an image, despite the aforementioned modifications, and to develop DP methods to protect the perceived information.

Singular Value Decomposition

In [Fan19b], Singular Value Decomposition (SVD) was considered to capture the perceptual information in input images. It is known that SVD can extract most of the geometric structure and characteristics of the image data. Prior work on perceptual image hashing methods [KVM04] based on SVD have been shown to robustly hash visually similar images, such as after compression, rotation, and cropping. The intuition of SVD is that any real or complex matrix A can be decomposed into a product of three matrices, i.e., $A = U\Sigma V^T$, where U and V are left and right singular vector matrices, Σ is a non-negative diagonal matrix, consisting of the singular values. Intuitively, the singular vectors in U and V, capture the geometric *features*

in an image, while the singular values in Σ can be interpreted as the *magnitudes* of the features.

Privacy in High-dimensional Spaces

As a strong attack model, we can assume an adversary who may have approximate knowledge about an input image. Specifically, the adversary knows the set of images that are visually similar to a given image (including the image itself), e.g., with the same singular matrices i.e., U and V, but different singular values, i.e., Σ. The adversarial goal is to infer the exact input image by observing the obfuscated image. To protect the private singular values, the study of [Fan19b] adopted a variant of differential privacy [CABP13] for up to k singular values, as follows.

Definition 11.4. [$\varepsilon \cdot d_k$-privacy] Suppose domains $\mathcal{X}, \mathcal{Z} \subset \mathbb{R}^k$ and d_k denotes the Euclidean metric in $\mathbb{R}^k$. A mechanism $K : \mathcal{X} \to \mathcal{Z}$ satisfies $\varepsilon \cdot d_k$-privacy, if and only if $\forall x, x' \in \mathcal{X}$,

$$K(x)(Z) \leq e^{\varepsilon \cdot d_k(x,x')} K(x')(Z) \quad \forall Z \in \mathcal{Z}. \tag{11.3}$$

This definition shows that the output of a mechanism should be "indistinguishable" to protect privacy, and the level of indistinguishability is proportional to the distance $d_k(x, x')$ between two inputs x and x'. For an adversary who observes the certain output Z, i.e., privacy-enhanced singular values, it is challenging to infer the exact input to the mechanism, i.e., real singular values. A sampling-based mechanism K in $\mathbb{R}^k$ was designed in [Fan19b] to satisfy Definition 11.4.

SVD-based Obfuscation (DP-SVD)

Figure 11.5 depicts the proposed framework for privacy-preserving image data sharing. An image often contains one or more regions-of-interest (ROIs), such as faces, objects, text, etc., where obfuscation is needed to protect privacy. Such ROIs can be detected automatically or annotated by data owners. The randomized obfuscation will be applied to the ROIs. The obfuscation step involves two components: transformation and sampling. A ROI will first be *transformed* to obtain the feature vector, i.e., singular values, and the vector will be truncated and go through the *sampling* step to achieve differential privacy guarantees; the sampled vector will be used in the *inverse* transform, resulting in the obfuscated ROI image.

Similar to standard obfuscations, applying DP-SVD incurs distortions in the image data. As can be seen in Figure 11.6, shapes in the image are distorted but high-level information is still perceptible after obfuscation. The intuition is that values close to the real singular values are sampled with higher probabilities, thanks to the relaxation of differential privacy (Definition 11.4). Evaluations that quantify the perceptual similarity between images will be discussed in the next section.

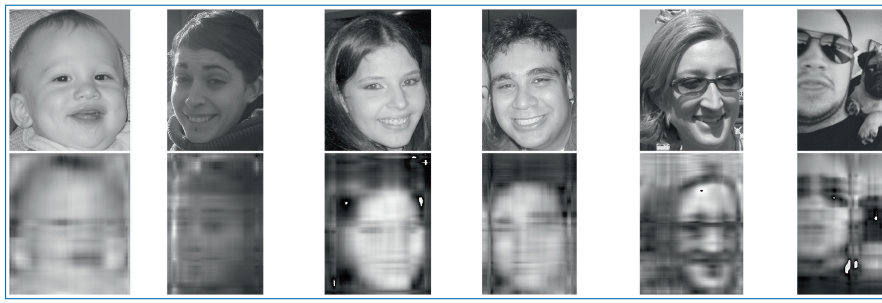


Figure 11.6. Example Images from the PIPA Dataset: row 1 – original images ; row 2 – images in Row 1 obfuscated by DP-SVD with $k = 4$ and $\varepsilon = 0.5$ [Fan19b].

11.2.3 Privacy in Videos

Essentially, a video consists of a sequence of frames, where each frame itself is an image. An natural extension of a DP definition for images in Sections 11.2.1 and 11.2.2 is to apply DP to every frame. However, it could be computationally expensive, especially given high "frames-per-second" (FPS) rates for current video cameras; furthermore, it results in a large privacy cost, i.e., the privacy guarantee degrades as the number of frames increases. It is thus important to quantify privacy for sensitive content, such as pedestrians and objects, which may appear in multiple frames in a video sequence.

Privacy for Input Videos

A recent work [WXH20] consider two videos as neighbors if they differ in at most one visual element, considering the element's appearance throughout the video sequence.

Definition 11.5. [Neighboring Videos] To protect sensitive visual elements in the video, two input videos V and V' that differ in any visual element γ in all frames are considered as two neighboring inputs. Note that V and V' have identical number of frames and background scene.

As can be seen, Definition 11.5 extends image DP to 3D, considering pixels belonging to a specific visual element (e.g., person or object) in all frames of the video. Based on the neighbor definition, the (ε, δ)-DP notion can be extended to videos.

Definition 11.6. [(ε, δ)-DP for Videos] A randomization algorithm $\mathcal{A}$ satisfies (ε, δ)-differential privacy if for every video V, we can divide the output space $range(\mathcal{A})$ into two sets $\mathbf{\Omega}_1$ and $\mathbf{\Omega}_2$ such that, (1) $\Pr[\mathcal{A}(V) \in \mathbf{\Omega}_1] \leq \delta$, and (2) for any of V's neighboring video V' and for all output $O \in \mathbf{\Omega}_2$, $e^{-\varepsilon} \leq \frac{\Pr[\mathcal{A}(V)]=O]}{\Pr[\mathcal{A}(V')]=O]} \leq e^{\varepsilon}$.


(a) Sampling video frame. (b) Synthetic video frame.

Figure 11.7. Example Sanitized Video Frames using MOT [Mil+16] Dataset: (a) video frame via pixel sampling [WXH20]; (b) video frame via synthesis [WHKV20].

The relaxation in Definition 11.6 is needed to account for the unique information contributed by a visual element in the input video. For instance, consider a mechanism $\mathcal{A}$ that randomly selects pixels from an input video. Furthermore, let γ be a vehicle in video V but not in V', i.e., $V' = V \setminus \gamma$. Observing any pixels of γ in $\mathcal{A}$'s output would allow an adversary to infer the presence of the vehicle in the input, as the probability of observing those pixels in $\mathcal{A}(V')$ is 0. The analysis of δ remains open in [WXH20]. A video frame produced by pixel sampling is shown in Figure 11.7(a). It can be seen that with weaker DP guarantees, visual elements, e.g., pedestrians with distinctive outfits, may still be identified.

Privacy for Occurrences

As illustrated in Figure 11.7(a), it is quite challenging to protect the presence of visual elements in videos. The privacy model may be relaxed to protecting the *occurrences* of each visual element. In other words, assume the presence of an visual element in the video is public (as in the *local DP* setting); the secret is which frames the visual element appears in. [WHKV20] proposed the following notion to achieve indistinguishability for the occurrences of visual elements (i.e., objects).

Definition 11.7. [ε-Object Indistinguishably] Suppose video V contains M frames and each object O_i is represented by a bit vector $B_i = \{b_i^k | k = 1, \cdots, M\}$, where b_i^k is set to 1 if O_i appears in the k-th frame and 0 otherwise. A randomization algorithm $\mathcal{A}$ satisfies ε-object indistinguishability if and only if for any two input objects $O_i, O_j \in \mathcal{O}$ in the input video V, and for any output object of $\mathcal{A}$ in the synthetic video V^* (denoted as y), we have

$$Pr[\mathcal{A}(O_i) = y] \leq e^{\varepsilon} Pr[\mathcal{A}(O_j) = y] \tag{11.4}$$

Based on Definition 11.7, it is intuitive that we can apply randomized response mechanisms, such as RAPPOR [EPK14], to the bit vectors of objects in order to satisfy Equation 11.4.

(a) Non-private Gaze Heatmap. (b) Private Gaze Heatmap.

Figure 11.8. Eye Gaze Heatmaps of an Individual User on a Web Page [LCFK21]: (a) raw data; (b) differentially private heatmap (ε=3).

Video Synthesis

Video frames can be sanitized by applying the privacy models in Definition 11.6 or Definition 11.7, and the mechanisms respectively. Figure 11.7 presents two sample output frames obtained respectively. As can be seen, sampling pixels allows for subsequent tasks, such as pedestrian detection. Video synthesis is less ambiguous, e.g., in terms of pedestrian counting, and does not disclosing identifying information by replacing pedestrians with icons. A natural question is: how do we evaluate the quality of output video frames? We will discuss that in the next section.

11.2.4 Adaptations to Eye Tracking

Eye-tracking applications capture large amounts of image and video data via webcams, wearable glasses, or mixed reality headsets. Eye gaze positions in a scene are used by eye-tracking applications to estimate what the user is viewing in order to prefetch contents or trigger events. An example gaze heatmap of a user on a web page is depicted in Figure 11.8(a). As can be seen, eye gaze data are essentially 2D positions in a given scene. It is thus intuitive that the geo-indistinguishability framework [ABCP13] may be adopted for eye-tracking applications [LCFK21].

Definition 11.8. [(ε, r)-geo-indistinguishability] A mechanism $\mathcal{M} : X \to Z$ is defined to be (ε, r)-geo-indistinguishable if and only if for all pairs of inputs $(x, x') \in X \times X$ such that $d(x, x') \leq r$,

$$\Pr[\mathcal{M}(x) \in S] \leq e^{\varepsilon \cdot d(x,x')} \Pr[\mathcal{M}(x') \in S], \forall S \subset Z \tag{11.5}$$

where $d(\cdot, \cdot)$ denotes the Euclidean metric.

Input x and x' give the corresponding eye gaze positions and the pair (x, x') can be considered as r-Euclidean neighbors if $d(x, x') \leq r$. As eye gaze streams are collected from each user, the authors of [LCFK21] adopted the w-event privacy model to achieve a privacy-utility tradeoff. A private heatmap with $\varepsilon = 3$ is depicted in Figure 11.8(b).

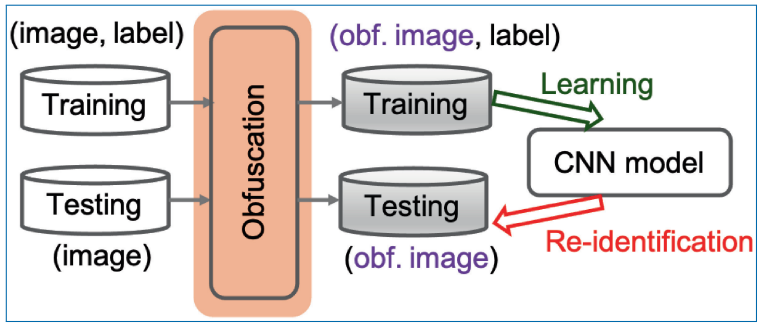


Figure 11.9. Adaptive Re-identification Evaluation for Image Obfuscation Methods: step 1 – both training and testing partitions go through the same type of obfuscation; step 2 – a CNN model is trained to predict identity labels on the training set; step 3 – the trained CNN model infers the identity labels of the test set.

11.3 Practical Considerations for DP Methods

In this section, we will discuss practical considerations for applying DP to image and video data. Specifically, we will look at measures of effective privacy protection and quantitative measures for output quality, which are important evaluation metrics for DP methods.

11.3.1 Effective Privacy Protection

The privacy guarantees of DP are well grounded theoretically. To facilitate the adoption of DP methods, it is important to understand the privacy protection achieved in practice, e.g., whether empirical privacy risks are mitigated by DP methods. Here we primarily illustrate how re-identification risks can be measured in practice and discuss adaptations for other risk measures.

Re-identification Risks

An important risk for visual data is that inference attacks can be launched against the sanitized data, despite the application of any obfuscation methods. To evaluate differentially private image obfuscation, we can carry out such inference attacks to understand the mitigation effects of DP. A CNN based re-identification attack was first proposed in [MSS16]. We adopt a similar attack below to evaluate the re-identification risks for DP-Pix and DP-Blur on two commonly used datasets.

Adaptive Re-identification Evaluation

An adaptive re-identification attack can be carried out to understand how much can be learned about a given obfuscation method, e.g., pixelization. The attack

Table 11.2. Accuracy (in %) of CNN Re-Identification Attacks [Fan19a].

Dataset	Random	Pix	DP-Pix ($b = 16$)			Blur	DP-Blur ($k = 99$)		
	–	$b = 16$	$\varepsilon = 0.1$	0.5	1	$k = 99$	$\varepsilon = 0.1$	0.5	1
AT&T	2.50	96.25	3.75	43.75	77.50	88.75	**1.25**	7.50	17.50
MNIST	**10.00**	52.13	16.41	21.51	22.95	76.35	11.35	11.75	13.43

was carried out to evaluate DP obfuscation methods in [Fan18; Fan19a], which is depicted in Figure 11.9. Here, "identity" refers to the user ID for a facial image, or the class label for hand-writings and objects. As can be seen, the same obfuscation method will be applied to image data in both training and testing partitions. In the context of DP obfuscation, we maintain the same privacy level (e.g., m and ε as in Definition 11.3) as well as other hyper-parameters (e.g., b for pixelization and k for Gaussian blur) when obfuscating training and testing images. That helps simulate a powerful attacker, who possesses knowledge about the obfuscation method. Subsequently, the obfuscated training set, which includes images and their labels, will be used to train a CNN-based deep model to predict labels. Note that the architecture of the CNN model could be adapted to each dataset. At the inference phase, the trained CNN model predicts labels for the obfuscated testing images. The accuracy of the prediction indicates the level of re-identification risks for the dataset and the obfuscation method.

Evaluation results for DP-Pix and DP-Blur are shown in Table 11.2, using the AT&T face database and the MNIST dataset. The "Random" column indicates the re-identification risks incurred by randomly guessing the label for each image. This column was populated according to the total number of classes in each dataset. Every other column shows the accuracy of re-identification for the corresponding obfuscation method as described above. As can be seen, the standard obfuscation methods, i.e., pixelization and Gaussian blur, will still allow an adversary to effectively learn to associate an obfuscated image with its label. The re-identification accuracy of faces is up to 96.25% and up to 76.35% for hand-written digits. These results may be surprising to some readers, especially when combined with the example images in Table 11.1. Again this showcases that images unrecognizable to human users may not be effectively *private*.

By introducing DP to image obfuscation, we can observe a reduction in re-identification accuracy, while the reduction depends on the parameters as well as the dataset. DP methods reduce face re-identification to as low as 1.25%, which is lower than random guessing, and MNIST re-identification to 11.35%. These results indicate that DP obfuscations provide stronger empirical privacy protection than standard obfuscation methods.

Re-identification for Eye Images and Videos

Similarly, features can be extracted from **videos** for re-identification evaluations. In [LCFK21], aggregate statistics of fixation/saccade features over several gaze video sessions were used to predict user's identity. The authors adopted a discriminant analysis classifier where the training and testing video sets correspond to the same privacy configuration. Re-identification for eye images requires specific methods, e.g., segmentation-based iris recognition [Gan+16]. Specifically, privacy-enhanced eye images will be compared to reference images and correct recognition rates [Joh+20] indicate the level of privacy risks. We refer interested readers to [RF21] for a comparative evaluation of DP methods on iris recognition risk mitigation.

Other Risk Measures

Attribute Inference Risks

Attributes of the person or object in image data are also subject to inference attacks. It is thus important to evaluate DP methods in mitigating attribute inference. For instance, multiple attributes, such as gender and smiling, can be inferred from facial images [CSVG18]. In [SHHB19], eye movement features were extracted to predict gender and to perform document type classification. [LCFK21] argued that scan-path features extracted from video gaze streams may distinguish users' psychophysiological traits. The framework in Figure 11.9 can be adapted for evaluating attribute inference risks. For instance, the choice of the model must correspond to the specific inference tasks. Another consideration is the nature of the dataset: the attack model can be either trained with clean data or DP-enhanced data.

Participation Inference Risks

Machine learning models trained with image data also leak information about the underlying training data. Although an adversary may not have direct access to the training images, membership inference (e.g., [SSSS17]) and model inversion (e.g., [FJR15]) can be carried out to predict whether an individual participates in the training set. A privacy researcher may adopt those attacks to evaluate the mitigation effects of DP methods in machine learning. An individual often contributes more than one image samples in the training set, e.g., in facial recognition applications (recall Figure 11.2). Therefore, model inversion may be applied to infer individual participation, with human users or machine classifiers to perform re-identification.

11.3.2 Quality Measures

It is also important to study the usefulness of DP methods in image and video analysis, in order to advance current research. Two types of measures are often adopted

to evaluate the output of DP methods: *task-based* measures for specific applications built on the DP outputs and *generic* measures that do not depend on any applications.

Generic Quality Measures

Absolute Errors and Perceptual Quality

The quality of privacy-enhanced images was measured by Mean Squared Error (MSE) and Structural Similarity (SSIM) in [Fan18]. In addition, we may also consider the peak signal-to-noise ratio (PSNR) measure, which represents the ratio between the maximum pixel value and the MSE. The quality of privacy-enhanced videos can be measured by adapting single image quality measures to video frames. MSE, PSNR, and SSIM measure the difference between the input image and the sanitized image, thus applicable in a wide of range settings. Among them, SSIM [WBSS04] is a widely used perceptual quality measure, which considers the perceived similarity in structural information in addition to luminance and contrast. One advantage of SSIM over absolute error based quality measures, is that an image derived by subtracting a certain value from every pixel in the input image would preserve high structural similarity, despite significant absolute errors. As an example, in Figure 11.10, we show that subtracting a constant pixel value leads to MSE=210 and almost perfect structural similarity, i.e., SSIM = 99%. Deep learning based image quality measures have been proposed recently, such as LPIPS [Zha+18] and SIFID [SDM19], which may also be applied to evaluate the DP obfuscated images.

Statistical Measures

Image and video data are often represented as color histograms, i.e., distributions of pixels across all possible colors. Therefore, distributional similarity measures can

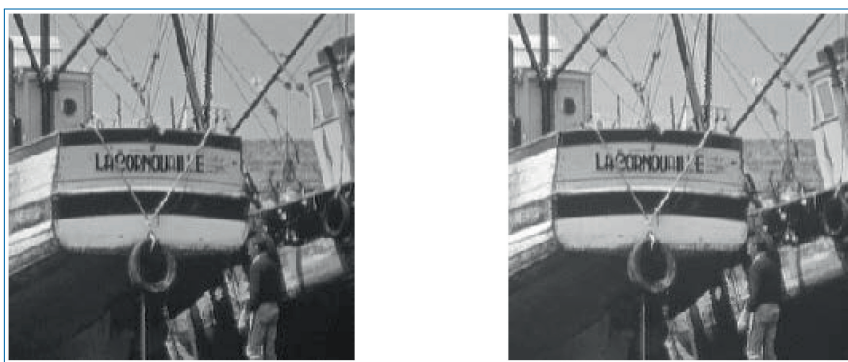


Figure 11.10. Comparison of "Boat" images: (a): original image; (b): mean-shifted image with MSE = 210 and SSIM = 99%. [WBSS04]

be adopted to compare the privacy-enhanced data with the input. For instance, [WXH20] adopted the Kullback–Leibler divergence for RGB pixel distributions between the input video and sanitized video. As another example, correlation coefficients were adopted to compare noisy gaze heatmaps to clean heatmaps [Liu+19]. When a set of images is generated with DP methods, e.g., in data synthesis [Fan20], measures are often adopted to quantify the realism and diversity of the generated dataset, e.g., via the Jensen-Shannon divergence and Inception Score.

Task-based Quality Measures

Specific tasks can be performed on the image and video data produced by DP methods. Performance measures for those tasks are thus suitable for evaluating the quality of DP outputs.

Image Analysis

Current DP methods aim to protect individual participation or identity while allowing for analysis tasks to be performed on the privacy-enhanced image data. For example, using the sanitized eye images, landmark detection and gaze estimation [PZBH18] can be performed. In addition, pupil detection was also evaluated on privacy-enhanced images in [Joh+20]. For facial images, we may consider attribute prediction tasks such as for gender and age, while protecting the identities. However, ensuring strong privacy and high utility for a single image may be challenging for DP based methods. We will discuss utility evaluation for machine learning based analysis shortly.

Video Analysis

In surveillance applications, video data is often analyzed to detect human [DT05] and objects [Yan+16]. Precision and recall measures can be adopted for detection tasks in privacy-enhanced video data. Furthermore, counting the total number of pedestrians or vehicles in a frame is important for anomaly detection applications [CZV08]. Such counting errors resulted from DP methods can be measured, e.g., using Mean Absolute Error. Moreover, tracking the appearance of a person or object in a video may also be of interest. For instance, [WXH20] measured the stay time of each pedestrian/vehicle, i.e., the number of video frames containing the pedestrian/vehicle; [WHKV20] measured how the trajectory of a pedestrian in the sanitized video frame sequence deviates from that of the original video.

Machine Learning

The performance of image models, such as for classification and segmentation, can be evaluated for DP-based machine learning methods, e.g., deep learning [Aba+16] and federated learning [Li+19]. Similarly, image models can be trained on privacy-enhanced image data, e.g., obtained via DP obfuscation or DP synthesis, and the

performance of those models may indicate the quality of the training data. For instance, in [TKP19], multiple classifiers for hand-written digits were trained using synthetic data generated by DP generative models; those classifiers were tested on real image data. Compared to the same classifiers trained on real data, close performance indicates that DP generative models could capture the real data distribution.

11.4 Concluding Remarks

So far, we have talked about how DP can be applied to protecting sensitive information in image and video data. We have also shown how to evaluate whether DP methods are successful, e.g., in achieving practical privacy protection and producing usefulness results. But the story does not stop there. The richness and complexity of image and video data, as well as the ubiquity of their applications, require collaborative research with experts beyond the DP community to address the privacy concerns.

User Perceptions of Privacy and Utility

The visual nature of images and videos dictates that privacy and utility are not only defined by mathematical equations, but also inseparable from end user perceptions in specific contexts. Recent studies [Li+17; Has+18] measured the viewer perceived privacy and utility for photos shared in the context of online social networks. Survey participants were asked to recognize persons, objects, and properties in obscured photos. For evaluating utility, participants were asked to rate the obscured photos in satisfaction, information sufficiency, visual appeals, etc. Users have been involved to evaluate privacy-preserving eye tracking via web cams [LCFK21]. Utility was quantified by scores users achieved in a game as well as self-reported enjoyment measures.

Deployment of DP Methods

As the local privacy mode is adopted in many image and video privacy methods, it is important to consider the feasibility to deploy those methods on user-owned devices. One consideration is the *computational overhead*. While DP methods based on aggregation (e.g., pixelization and blurring, gaze heatmaps) may not inflict significant overhead [SRF22], some problems/settings may be more challenging, e.g., private sampling in high-dimensional spaces and sanitizing videos with a large number of frames. Another consideration is the *integration* with devices and exiting platforms. We may need to consider the compatibility of DP methods with different cameras, imaging systems, and image and video analysis applications.

Acknowledgements

Liyue Fan is an Assistant Professor of Computer Science at UNC Charlotte. This work is supported in part by the National Science Foundation CNS-1949217, CNS-1951430, CNS-2144684, and the University of North Carolina. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the sponsors, such as the National Science Foundation.

References

[Aba+16] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang. "Deep learning with differential privacy". In: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. ACM. 2016, pp. 308–318 (cit. on pp. 391, 405).

[ABCP13] M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi. "Geo-indistinguishability: Differential privacy for location-based systems". In: Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security. 2013, pp. 901–914 (cit. on p. 400).

[Ano98] Anonymous. "To Reveal or Not to Reveal: A Theoretical Model of Anonymous Communication". In: Communication Theory 8.4 (1998), pp. 381–407. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1468-2885.1998.tb00226.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-2885.1998.tb00226.x (cit. on p. 388).

[BCCW19] F. Boemer, A. Costache, R. Cammarota, and C. Wierzynski. "NGraph-HE2: A High-Throughput Framework for Neural Network Inference on Encrypted Data". In: Proceedings of the 7th ACM Workshop on Encrypted Computing Applied Homomorphic Cryptography. WAHC-19. London, United Kingdom: Association for Computing Machinery, 2019, pp. 45–56. ISBN: 9781450368292. URL: https://doi.org/10.1145/3338469.3358944 (cit. on p. 389).

[CABP13] K. Chatzikokolakis, M. E. Andrés, N. E. Bordenabe, and C. Palamidessi. "Broadening the Scope of Differential Privacy Using Metrics". In: Privacy Enhancing Technologies. Ed. by E. De

Cristofaro and M. Wright. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 82–102. ISBN: 978-3-642-39077-7 (cit. on p. 397).

[CSVG18] S. Chhabra, R. Singh, M. Vatsa, and G. Gupta. "Anonymizing k-facial attributes via adversarial perturbations". In: arXiv preprint arXiv:1805.09380 (2018) (cit. on p. 403).

[CZV08] A. B. Chan, Zhang-Sheng John Liang, and N. Vasconcelos. "Privacy preserving crowd monitoring: Counting people without people models or tracking". In: 2008 IEEE Conference on Computer Vision and Pattern Recognition. 2008, pp. 1–7 (cit. on p. 405).

[DER15] A. Dantcheva, P. Elia, and A. Ross. "What else does your biometric data reveal? A survey on soft biometrics". In: IEEE Transactions on Information Forensics and Security 11.3 (2015), pp. 441–467 (cit. on p. 390).

[DR+14] C. Dwork, A. Roth, et al. "The algorithmic foundations of differential privacy." In: Foundations and Trends in Theoretical Computer Science 9.3-4 (2014), pp. 211–407 (cit. on pp. 392, 393, 395).

[DT05] N. Dalal and B. Triggs. "Histograms of oriented gradients for human detection". In: 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05). Vol. 1. Ieee. 2005, pp. 886–893 (cit. on p. 405).

[Dug15] M. Duggan. "Mobile Messaging and Social Media – 2015". In: Pew Research Center (Aug. 2015) (cit. on p. 387).

[EPK14] Ú. Erlingsson, V. Pihur, and A. Korolova. "RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response". In: Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security. CCS '14. Scottsdale, Arizona, USA: Association for Computing Machinery, 2014, pp. 1054–1067. ISBN: 9781450329576. URL: https://doi.org/10.1145/2660267.2660348 (cit. on p. 399).

[Fan18] L. Fan. "Image pixelization with differential privacy". In: IFIP Annual Conference on Data and Applications Security and Privacy. Springer. 2018, pp. 148–162 (cit. on pp. 394, 395, 402, 404).

[Fan19a] L. Fan. "Differential Privacy for Image Publication". In: Theory and Practice of Differential Privacy Workshop. 2019 (cit. on pp. 395, 402).

[Fan19b] L. Fan. "Practical Image Obfuscation with Provable Privacy". In: 2019 IEEE International Conference on Multimedia and Expo (ICME). 2019, pp. 784–789 (cit. on pp. 396–398).

[Fan20] L. Fan. "A survey of differentially private generative adversarial networks". In: The AAAI Workshop on Privacy-Preserving Artificial Intelligence. 2020 (cit. on p. 405).

[FJR15] M. Fredrikson, S. Jha, and T. Ristenpart. "Model inversion attacks that exploit confidence information and basic countermeasures". In: Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security. 2015, pp. 1322–1333 (cit. on pp. 390, 391, 403).

[Gan+16] A. Gangwar, A. Joshi, A. Singh, F. Alonso-Fernandez, and J. Bigun. "IrisSeg: A fast and robust iris segmentation framework for non-ideal iris images". In: 2016 International Conference on Biometrics (ICB). 2016, pp. 1–8 (cit. on p. 403).

[Has+18] R. Hasan, E. Hassan, Y. Li, K. Caine, D. J. Crandall, R. Hoyle, and A. Kapadia. "Viewer experience of obscuring scene elements in photos to enhance privacy". In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018, pp. 1–13 (cit. on pp. 388, 406).

[HLP12] C. Y. Hsu, C. S. Lu, and S. C. Pei. "Image Feature Extraction in Encrypted Domain With Privacy-Preserving SIFT". In: IEEE Transactions on Image Processing 21.11 (Nov. 2012), pp. 4593–4607. ISSN: 1057-7149 (cit. on p. 389).

[Hoy+20] R. Hoyle, L. Stark, Q. Ismail, D. Crandall, A. Kapadia, and D. Anthony. "Privacy norms and preferences for photos Posted Online". In: ACM Transactions on Computer-Human Interaction (TOCHI) 27.4 (2020), pp. 1–27 (cit. on p. 388).

[HRSF20] Y. He, S. Rahimian, B. Schiele, and M. Fritz. "Segmentations-leak: Membership inference attacks and defenses in semantic image segmentation". In: European Conference on Computer Vision. Springer. 2020, pp. 519–535 (cit. on p. 391).

[HZSS16] S. Hill, Z. Zhou, L. Saul, and H. Shacham. "On the (in) effectiveness of mosaicing and blurring as tools for document redaction". In: Proceedings on Privacy Enhancing Technologies 2016.4 (2016), pp. 403–417 (cit. on p. 390).

[Joh+20] B. John, A. Liu, L. Xia, S. Koppal, and E. Jain. "Let It Snow: Adding pixel noise to protect the user's identity". In: ACM Symposium on Eye Tracking Research and Applications. 2020, pp. 1–3 (cit. on pp. 393, 403, 405).

[Kno13] J. Knott. Video Surveillance Recordings Create 413 Petabytes of Data Every Day, http : / / www . cepro . com / article / video _ surveillance_recordings_create_413_petabytes_of_data_every_day. 2013 (cit. on p. 388).

[KVM04] S. S. Kozat, R. Venkatesan, and M. K. Mihcak. "Robust perceptual image hashing via matrix invariants". In: Image Processing, 2004. ICIP '04. 2004 International Conference on. Vol. 5. Oct. 2004, 3443–3446 Vol. 5 (cit. on p. 396).

[LCFK21] J. Li, A. R. Chowdhury, K. Fawaz, and Y. Kim. "Kalεido: Real-Time Privacy Control for Eye-Tracking Systems". In: 30th USENIX Security Symposium (USENIX Security 21). Vancouver, B.C.: USENIX Association, Aug. 2021. URL: https://www.usenix.org/conference/usenixsecurity21/presentation/li-jingjie (cit. on pp. 400, 403, 406).

[Lea+15] L. Leal-Taixé, A. Milan, I. Reid, S. Roth, and K. Schindler. "Motchallenge 2015: Towards a benchmark for multi-target tracking". In: arXiv preprint arXiv:1504.01942 (2015) (cit. on p. 394).

[Li+17] Y. Li, N. Vishwamitra, B. P. Knijnenburg, H. Hu, and K. Caine. "Effectiveness and users' experience of obfuscation as a privacy-enhancing technology for sharing photos". In: Proceedings of the ACM on Human-Computer Interaction 1.CSCW (2017), pp. 1–24 (cit. on pp. 388, 406).

[Li+19] W. Li, F. Milletarì, D. Xu, N. Rieke, J. Hancox, W. Zhu, M. Baust, Y. Cheng, S. Ourselin, M. J. Cardoso, et al. "Privacy-preserving federated brain tumour segmentation". In: International Workshop on Machine Learning in Medical Imaging. Springer. 2019, pp. 133–141 (cit. on pp. 391, 405).

[Liu+19] A. Liu, L. Xia, A. Duchowski, R. Bailey, K. Holmqvist, and E. Jain. "Differential privacy for eye-tracking data". In: Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications. 2019, pp. 1–10 (cit. on p. 405).

[LTKC18] Y. Li, W. Troutman, B. P. Knijnenburg, and K. Caine. "Human perceptions of sensitive content in photos". In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2018, pp. 1590–1596 (cit. on p. 388).

[Mar01] G. Marx. Identity and anonymity: Some conceptual distinctions and issues for research. Princeton University Press, 2001 (cit. on p. 388).

[Mil+16] A. Milan, L. Leal-Taixé, I. D. Reid, S. Roth, and K. Schindler. "MOT16: A Benchmark for Multi-Object Tracking". In: CoRR abs/1603.00831 (2016). arXiv: 1603.00831. URL: http://arxiv.org/abs/1603.00831 (cit. on p. 399).

[MRR18] V. Mirjalili, S. Raschka, and A. Ross. "Gender Privacy: An Ensemble of Semi Adversarial Networks for Confounding Arbitrary Gender Classifiers". In: 2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS). 2018, pp. 1–10 (cit. on p. 389).

[MSS16] R. McPherson, R. Shokri, and V. Shmatikov. "Defeating Image Obfuscation with Deep Learning". In: CoRR abs/1609.00408 (2016) (cit. on pp. 390, 401).

[NBB19] J. F. Nettrour, M. B. Burch, and B. S. Bal. "Patients, pictures, and privacy: managing clinical photographs in the smartphone era". In: Arthroplasty today 5.1 (2019), pp. 57–60 (cit. on p. 388).

[NSM05] E. M. Newton, L. Sweeney, and B. Malin. "Preserving privacy by de-identifying face images". In: IEEE transactions on Knowledge and Data Engineering 17.2 (2005), pp. 232–243 (cit. on p. 389).

[OR15] A. Othman and A. Ross. "Privacy of Facial Soft Biometrics: Suppressing Gender But Retaining Identity". In: Computer Vision - ECCV 2014 Workshops. Ed. by L. Agapito, M. M. Bronstein, and C. Rother. Cham: Springer International Publishing, 2015, pp. 682–696. ISBN: 978-3-319-16181-5 (cit. on p. 389).

[Pap+16] N. Papernot, M. Abadi, U. Erlingsson, I. Goodfellow, and K. Talwar. "Semi-supervised knowledge transfer for deep learning from private training data". In: arXiv preprint arXiv:1610.05755 (2016) (cit. on p. 391).

[PM92] W. Pennebaker and J. Mitchell. JPEG: Still Image Data Compression Standard. Chapman & Hall digital multimedia standards series. Springer US, 1992. ISBN: 9780442012724. URL: https://books.google.com/books?id=AepB%5C_PZ%5C_WMkC (cit. on p. 396).

[PZBH18] S. Park, X. Zhang, A. Bulling, and O. Hilliges. "Learning to Find Eye Region Landmarks for Remote Gaze Estimation in Unconstrained Settings". In: Proceedings of the 2018 ACM Symposium on Eye Tracking Research Applications. ETRA '18. Warsaw, Poland: Association for Computing Machinery, 2018. ISBN: 9781450357067. URL: https://doi.org/10.1145/3204493.3204545 (cit. on p. 405).

[RF21] D. Reilly and L. Fan. "A comparative evaluation of differentially private image obfuscation". In: 2021 Third IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA). IEEE. 2021, pp. 80–89 (cit. on p. 403).

[RGO13] M.-R. Ra, R. Govindan, and A. Ortega. "P3: Toward Privacy-Preserving Photo Sharing". In: Presented as part of the 10th USENIX Symposium on Networked Systems Design and Implementation (NSDI 13). Lombard, IL: USENIX, 2013, pp. 515–528. ISBN: 978-1-931971-00-3. URL: https://www.usenix.org/conference/nsdi13/technical-sessions/presentation/ra (cit. on pp. 389, 390).

[RGRB19] A. Rozsa, M. Günther, E. M. Rudd, and T. E. Boult. "Facial attributes: Accuracy and adversarial robustness". In: Pattern Recognition Letters 124 (2019), pp. 100–108 (cit. on p. 389).

[SCB17] A. Squicciarini, C. Caragea, and R. Balakavi. "Toward automated online photo privacy". In: ACM Transactions on the Web (TWEB) 11.1 (2017), pp. 1–29 (cit. on p. 388).

[Sco04] C. R. Scott. "Benefits and drawbacks of anonymous online communication: Legal challenges and communicative recommendations". In: Free speech yearbook 41.1 (2004), pp. 127–141 (cit. on p. 388).

[SDM19] T. R. Shaham, T. Dekel, and T. Michaeli. "Singan: Learning a generative model from a single natural image". In: Proceedings of the IEEE/CVF international conference on computer vision. 2019, pp. 4570–4580 (cit. on p. 404).

[SHHB19] J. Steil, I. Hagestedt, M. X. Huang, and A. Bulling. "Privacy-aware eye tracking using differential privacy". In: Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications. 2019, pp. 1–9 (cit. on p. 403).

[Smi17] A. Smith. Record shares of Americans have smartphones, home broadband. https://www.pewresearch.org/fact-tank/2017/01/12/evolution-of-technology/. Accessed: 2021-02-28. Jan. 2017 (cit. on p. 387).

[SP17] R. Stevens and I. Pudney. Blur select faces with the updated Blur Faces tool. Ed. by youtube-eng.googleblog.com. [Online; posted 21-August-2017]. Aug. 2017. URL: https://youtube-eng.googleblog.com/2017/08/blur-select-faces-with-updated-blur.html (cit. on pp. 389, 390).

[SRF22] M. U. Saleem, D. Reilly, and L. Fan. "DP-Shield: Face Obfuscation with Differential Privacy". In: Proceedings of the 25th International Conference on Extending Database Technology, EDBT 2022, Edinburgh, UK, March 29 - April 1, 2022. Ed. by J. Stoyanovich, J. Teubner, P. Guagliardo, M. Nikolic, A. Pieris, J. M'uhlig, F. 'Ozcan, S. Schelter, H. V. Jagadish, and M. Zhang. OpenProceedings.org, 2022, 2:578–2:581. URL: https://doi.org/10.48786/edbt.2022.55 (cit. on p. 406).

[SSSS17] R. Shokri, M. Stronati, C. Song, and V. Shmatikov. "Membership inference attacks against machine learning models". In: 2017 IEEE symposium on security and privacy (SP). IEEE. 2017, pp. 3–18 (cit. on pp. 391, 403).

[TKP19] R. Torkzadehmahani, P. Kairouz, and B. Paten. "DP-CGAN: Differentially Private Synthetic Data and Label Generation". In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. June 2019 (cit. on pp. 391, 406).

[WBSS04] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. "Image quality assessment: from error visibility to structural similarity". In: IEEE Transactions on Image Processing 13.4 (Apr. 2004), pp. 600–612. ISSN: 1057-7149 (cit. on p. 404).

[WHKV20] H. Wang, Y. Hong, Y. Kong, and J. Vaidya. "Publishing video data with indistinguishable objects". In: 23rd International Conference on Extending Database Technology, EDBT 2020. OpenProceedings.org. 2020, pp. 323–334 (cit. on pp. 399, 405).

[WK18] Y. Wang and M. Kosinski. "Deep neural networks are more accurate than humans at detecting sexual orientation from facial images." In: Journal of personality and social psychology 114.2 (2018), p. 246 (cit. on p. 390).

[WR14] T. Winkler and B. Rinner. "Security and privacy protection in visual sensor networks: A survey". In: ACM Computing Surveys (CSUR) 47.1 (2014), pp. 1–42 (cit. on p. 388).

[WXH20] H. Wang, S. Xie, and Y. Hong. "VideoDP: A Flexible Platform for Video Analytics with Differential Privacy". In: Proceedings on Privacy Enhancing Technologies 2020.4 (2020), pp. 277–296 (cit. on pp. 398, 399, 405).

[Xia+16] Z. Xia, X. Wang, L. Zhang, Z. Qin, X. Sun, and K. Ren. "A Privacy-Preserving and Copy-Deterrence Content-Based Image Retrieval Scheme in Cloud Computing". In: IEEE Transactions on Information Forensics and Security 11.11 (Nov. 2016), pp. 2594–2608. ISSN: 1556-6013 (cit. on p. 389).

[Xie+18] L. Xie, K. Lin, S. Wang, F. Wang, and J. Zhou. "Differentially private generative adversarial network". In: arXiv preprint arXiv:1802.06739 (2018) (cit. on p. 391).

[Yan+16] J. Yang, B. Price, S. Cohen, H. Lee, and M.-H. Yang. "Object contour detection with a fully convolutional encoder-decoder network". In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016, pp. 193–202 (cit. on p. 405).

[Zha+18] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang. "The unreasonable effectiveness of deep features as a perceptual metric". In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2018, pp. 586–595 (cit. on p. 404).

[Zha+20] Y. Zhang, R. Jia, H. Pei, W. Wang, B. Li, and D. Song. "The Secret Revealer: Generative Model-Inversion Attacks Against Deep Neural Networks". In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 2020 (cit. on p. 390).